

Корсакова Елена Анатольевна
кандидат технических наук,
магистрант кафедры фундаментальной
и прикладной лингвистики
Уральского гуманитарного института, УрФУ
им. первого Президента России Б.Н. Ельцина,
Екатеринбург, Россия
E-mail: e.a.korsakova@urfu.ru
Мухин Михаил Юрьевич
доктор филологических наук, доцент,
профессор кафедры фундаментальной
и прикладной лингвистики
Уральского гуманитарного института УрФУ
им. первого Президента России Б.Н. Ельцина,
Екатеринбург, Россия

ПРЕПОДАВАНИЕ АКАДЕМИЧЕСКОГО АНГЛИЙСКОГО СТУДЕНТАМ-ОПТИКАМ НА ОСНОВЕ КОРПУСНОГО ПОДХОДА: ВОЗМОЖНОСТИ ЦИФРОВОЙ ЛИНГВИСТИКИ

УДК 378.4

Аннотация. В статье проведен обзор использования инструментов цифровой лингвистики в целях разработки высокоэффективных образовательных продуктов в сфере академического английского и английского для конкретных целей. Выявлено три основных пути внедрения технологий цифровой лингвистики в преподавание английского языка: 1. составление учебно-методических материалов на основе корпусного подхода; 2. проведение студентами непосредственных исследований текстовых корпусов с помощью цифровых инструментов; 3. создание инновационных элементов визуализации языка – коллокационных сетей. Обоснована актуальность и показаны возможности развития данных направлений для преподавания английского языка в области оптики.

Ключевые слова: английский язык для конкретных целей, #LancsBox, корпусный подход, цифровая лингвистика, высокочастотные коллокации, коллокационные сети, преподавание английского в оптике

CORPUS-BASED ACADEMIC ENGLISH TEACHING IN OPTICS: OPPORTUNITIES PROVIDED BY DIGITAL LINGUISTICS

Dr. E.A. Korsakova*, Master's student

Prof. M.Yu. Mukhin

Department of Fundamental and Applied Linguistics, and Textology
Ural Federal University named after the first President of Russia B.N. Yeltsin,
Russia, Yekaterinburg, *E-mail: e.a.korsakova@urfu.ru

Abstract. This article provides an overview of implementing digital linguistics tools for the development of highly effective educational products in the field of Academic English and English for Specific Purposes. Three main ways of such an implementation were revealed: the corpus-based development of instructional materials; the investigation of text corpora by students via digital

tools; the building of innovative language visualization elements – collocation networks. We also show the relevance of this topic and the opportunities provided by digital linguistics for teaching English in the field of optics. We expect this approach, while boosting educational innovation, will allow improving the quality of teaching English in optics, positively affecting the performance of students.

Keywords: ESP, #LancsBox, corpus-based approach, computational linguistics, high-frequency collocations, collocation networks, English teaching in optics

Введение

В диджитал-эпоху цифровые инструменты входят во все сферы жизни. Это в полной мере касается и высшего образования. Трансформация и повышение качества происходит на каждом этапе – от разработки до реализации образовательных продуктов. Если говорить о сфере преподавания английского языка, то в ней все чаще используют корпусный подход, который обеспечивается средствами цифровой лингвистики.

Цифровая лингвистика – это наука об управлении цифровыми данными для целей и задач лингвистики, включая цифровое хранение, распространение, представление и обработку лингвистических данных. Причем обработке информации уделяется особое внимание в этой науке. В последние десятилетия в большинстве своем для этого используется корпусный подход, который подразумевает использование данных текстовых корпусов для изучения теории или гипотезы – их проверки, опровержения или уточнения [1; 2]. К настоящему времени разработано достаточно много программных пакетов для цифровой лингвистики, позволяющих проводить анализ текстовых корпусов: Range, AntConc, Compleat Lexical Tutor и LancsBox [3]. Последний программный пакет имеет ряд преимуществ перед своими аналогами. #LancsBox v.5.1.2. – разработка ученых из Ланкастерского университета – бесплатное программное обеспечение нового поколения для анализа языковых данных и корпусов, дающее возможность проводить автоматическое тегирование (аннотирование) текста и использовать новейшие инструменты статистического анализа, а также инновационные возможности визуализации [4].

Использование таких передовых технологий в преподавании иностранных языков позволяет повысить продуктивность преподавателей и успеваемость студентов. На уровне лингвистических единиц это проявляется в поиске и изучении высокочастотных слов и коллокаций (т. е. словосочетаний, имеющих признаки синтаксически и семантически целостных единиц, в которых выбор одного из компонентов осуществляется по смыслу, а выбор второго зависит от выбора первого). В научной литературе приведены многочисленные примеры эффективности этого подхода. Есть успешные case study по созданию практических занятий, основанных на корпусном подходе, в курсе английского для студентов-физиков (родной язык – итальянский) [5]; по преподаванию студентам юридических специальностей специализированной английской лексики, отобранной на основе корпусного подхода, (родной язык – персидский) [6]; по имплементации инструментов цифровой лингвистики в обучение академическому переводу аспирантов медицинских направлений (родной язык – испанский) [7]. Отмечается, что цифровой корпусный подход значительно повышает энтузиазм студентов к самообучению, их успеваемость растет [8].

Несмотря на успехи, исследований, посвященных изучению корпусов научных текстов по оптике, крайне мало [9; 10]. Совсем нет исследований специальных коллокаций в оптике, а, следовательно, нет и учебных материалов в этой области, разработанных с применением передовых методик. Найден всего один коммерчески доступный корпус специальных текстов по оптике, однако только на французском языке [11]. А ведь оптика является динамично развивающейся областью науки и техники, и подготовка высококвалифицированных конкурентоспособных специалистов в этой сфере входит в приоритетные задачи профильных вузов. Чтобы восполнить этот пробел, было принято решение инициировать разработку учебных материалов по английскому языку для студентов оптических специальностей с использованием технологий цифровой лингвистики.

Цель данной статьи – провести обзор примеров использования инструментов цифровой лингвистики в целях разработки высокоэффективных образовательных продуктов в сфере академического английского (Academic English, AE) и английского для специальных целей (English for specific purposes, ESP), и на основе проведенного обзора обосновать актуальность тематики и выявить возможности, которые предоставляют данные инструменты для преподавания английского языка в области оптики.

Методы и материалы

Рассмотрение темы осуществлялось на материале релевантных научных статей рецензируемых журналов из списка библиографической и реферативной базы данных Scopus. Был проведен обзор статей по следующей тематике: 1) корпусные исследования научных текстов по оптике на английском языке; 2) исследования специальных коллокаций в области оптики на английском языке; 3) работы по составлению частотных списков слов и коллокаций для целей изучения ESL (English as a Second Language) в различных академических областях; 4) примеры внедрения корпусного подхода, частотных списков слов/коллокаций и учебных материалов на их основе в процесс обучения ESP.

Апробация инструментов цифровой лингвистики проводилась с помощью программного пакета #LancsBox на корпусе из 66 научных статей по оптике, содержащем 321279 токенов, 35540 лемм. Были использованы следующие преднастройки: коллокационное окно – 3L 3R (т.е. три слова слева и три слова справа от интересующего слова); количественная ассоциативная мера – LogDice (отношение гармонического среднего от двух пропорций, выражающего тенденцию двух слов к совместному появлению в корпусе, к частоте появления этих слов в корпусе [12]); поиск ключевых слов (узлов) проводился без лемматизации.

Результаты и обсуждение

В ходе анализа научных статей удалось выявить несколько путей внедрения технологий цифровой лингвистики в преподавание английского языка:

1. использование программных пакетов для составления учебных корпусов, пособий, частотных списков слов и коллокаций, учебных словарей и так далее, т. е. учебно-методических материалов;
2. проведение студентами «полевых исследований» текстовых корпусов;
3. создание коллокационных сетей – элементов визуализации, которые помогают студентам лучше запомнить различные лексические и грамматические структуры и связи между ними.

Первый путь внедрения возможностей цифровой лингвистики в преподавание иностранных языков является самым масштабным из всех перечисленных. Это косвенно подтверждается большим количеством публикаций по данной тематике. В 2019 году опубликовано исследование по отбору (на основе комбинации экспертной оценки и корпусного подхода) более 400 групп этимологически родственных слов специализированной академической лексики по 11 поддисциплинам естествознания [13]. Есть ряд исследований по изучению специальной лексики с помощью корпусного подхода, доказывающих успешность этого метода: создание сбалансированного списка технической лексики исламских религиозных наук, таких как исламское право, юриспруденция, исламская философия и теология [14], создание частотного списка академической лексики на русском языке с использованием для валидации перекрестной проверки и охвата вне выборки [15]. В области оптики подобных исследований не проводилось, насколько позволяет судить проведенный обзор.

В 2017 году опубликована статья [3]. В частности, в ней описывается, как разработать учебные материалы для обучения лексике L2 (second language – изучаемый язык) с использованием корпусных методик. Авторы обсуждают три

типа лексики, которые могут быть использованы в курсе английского языка для конкретных целей (технические, политехнические и нетехнические слова), и объясняют принципы, используемые для управления процессом выбора словарного запаса (релевантность для использования в роли ключевых слов, частота слов и их распределение в материалах). В статье Е. Корино 2019 года приведен успешный *case study* по созданию практических занятий в курсе английского в паре с итальянским (L1 – first language, или родной язык) для студентов-физиков с разработкой методологии, основанной на анализе корпусов, и привлечением инструментальных средств обучения [5]. Более близких исследований к теме ESP «оптика» в языковой паре русский-английский найдено не было.

В результате обзора публикаций удалось найти всего две статьи, посвященные корпусным исследованиям по оптике. В статье 2014 года анализируются жанровые особенности корпуса, составленного из публикаций «Journal of Optics», а именно частотность n-грамм, без деления на общую лексику академического письма и специальную лексику [9]. В результате, в статье достаточно мало представлены термины из области оптики. В статье 2016 года, на более крупном корпусе статей из «Journal of Optics» исследуются специфические обороты речи, используемые в каждом из структурных разделов статей (1. предпосылки исследования, 2. цель исследования, 3. методы, материалы, предметы и процедуры, 4. результаты и выводы, 5. обсуждение и заключение), предлагая результаты анализа в виде списков слов, n-грамм, кластеров и слов в контексте [10]. Что касается корпусных исследований непосредственно специальных коллокаций, относящихся к предметной области оптики, то в открытом доступе данных о них не было найдено, равно как и нет корпусов текстов из области оптики на английском языке.

В данном случае, программа #LancsBox будет полезной, поскольку позволяет формировать собственные текстовые корпуса, работая с различными форматами (txt, doc(x), pdf, xml и т.д.), и проводить их автоматическое

тегирование по частям речи, таким образом, что они сразу становятся годны для анализа. После этого их можно легко анализировать и составлять различные виды учебно-методических материалов ESP по оптике: учебных корпусов, пособий, частотных списков слов и коллокаций, учебных словарей и т.п. Так, в таблице представлен пример списка высокочастотных коллокаций из рассматриваемой предметной области. В этом списке коллокация «silver halide» является самой высокочастотной, что отражается наибольшей относительной частотой появления этого словосочетания в корпусе – 9,464125 (т.е. встречается ~ 9,5 раз на каждые 10 000 слов корпуса) (Таблица 1).

Таблица 1 – Пример списка высокочастотных коллокаций из области оптики

Коллокация L2	Эквивалент L1	Относит. частота
silver halide	галогенидсеребряный	9,464125
optical fiber	оптическое волокно	7,160358
refractive index	показатель преломления	5,136778
fiber bundle	волоконная сборка	4,763195
IR fiber	ИК волокно	4,389611
chalcogenide glass	халькогенидное стекло	4,202819
optical losses	оптические потери	3,922631
fiber optics	волоконная оптика	3,082067
thermal imaging	тепловая визуализация	2,801879
CO ₂ laser	СО ₂ лазер	2,739615
spectral range	спектральный диапазон	2,459427
single mode	одномодовый	2,272635
ordered bundle	упорядоченная сборка	2,116975
core diameter	диаметр сердцевины	1,930183
absorption coefficient	коэффициент поглощения	1,899051
thermal image	тепловая картинка	1,712259
numerical aperture	числовая апертура	1,556600
distal end	дистальный конец	1,556600
photonic crystal	фотонный кристалл	1,525468

В рамках второго пути внедрения технологий цифровой лингвистики в преподавание иностранного языка программные пакеты позволяют студентам изучать язык более эффективным способом, чем при традиционном подходе –

переводе определенного количества статей L2→L1. В статье 2020 года описан удачный *case study* по имплементации непосредственно цифровых инструментов корпусного подхода в обучение академическому переводу (испанский-английский) аспирантов медицинских направлений [7]. 40 аспирантов одного из мадридских университетов участвовали в составлении собственных текстовых корпусов и были обучены проведению анализа данных корпусов для решения своих тематических, терминологических и фразеологических задач. Результаты эксперимента показали, что студенты полностью удовлетворены использованием корпусного подхода. Судя по имеющимся публикациям, применительно к оптике подобных экспериментов не проводилось.

В это направление хорошо вписывается инструмент #LancsBox KWIC (key word in context, или «ключевое слово в контексте»). Он позволяет: находить частоту употребления слова или фразы в корпусе; анализировать частоту использования различных частей речи, таких как существительные, глаголы, прилагательные; используя «умный поиск», находить сложные лингвистические структуры, например, такие как пассивный залог. Результаты выдаются в виде списка конкордансов, в которых ключевое слово дается в контексте. На рисунке 1 представлен список конкордансов для ключевого слова «optics» (оптика). Коллокационное окно 3L 3R позволяет увидеть ближайшее окружение узла «optics». В абсолютном выражении частота этого слова в корпусе составляет 418, относительная частота этого слова в корпусе равна 13,01 на каждые 10 000 слов корпуса, а дисперсия – 66 из 66, т. е. слово «optics» встречается во всех статьях корпуса.

Index	File	Left	Node	Right
2	10.1038@npl		optics	associated with phase
5	Advancedmct	synergy with IR-fiber	optics	Galina V. Chekanova,
7	Advancedmct	based on IR-fiber	optics	is required photoconductors
9	Advancedmct	with innovative PIR-fiber	optics	opens a unique
14	artyushenko2	of specialty fibre	optics	includes specialty silica
16	artyushenko2	types of fiber	optics	for 0.2-4µm
19	artyushenko2	168-171 Mid-IR fibre	optics	spectroscopy in the
20	artyushenko2	in IR fibre	optics	enables to expand
22	artyushenko2	However, additional focusing	optics	complicates the probe
23	artyushenko2	practice of fibre	optics	for the same
25	artyushenko2	2004. Mid-IR fibre	optics	spectroscopy in the
42	chandraskh	occurs in fibre	optics	when a short
53	chen2016.pdt	al. Free-space coupling	optics	for multicore fibers.
59	Churbanov 1-	glasses for infrared	optics	and thermoelectricity". He
61	Churbanov 1-	glasses for infrared	optics	and thermoelectricity". Catherine
63	gannot2004.p	consisted of IR	optics	(e.g. ZnSe lenses,
75	Gopal gal201	There were no	optics	between the object
87	karbasi2014 i	and no additional	optics	or pre- and
104	lu2016.pdf	using appropriate mirror	optics	(Fig. 1) such
105	lu2016.pdf	xyz-positioners of the	optics	i.e. mirrors or
110	Matsuura non	codes: (170.3990) Medical	optics	instrumentation; (060.2390) Fiber
124	Mid-Infrared F	where the focusing	optics	is fixed together
129	Mid-Infrared F	on polycrystalline fiber	optics	for mid IR-range.
143	naeh2016+.p	majority of adaptive	optics	sys-tems are
148	naeh2016+.p	wavefront using adaptive	optics	techniques. The literature
155	optical losses	codes: (060.0060) Fiber	optics	and optical communications;
158	optical losses	in a geometrical	optics	approximation: T i
171	paiss1991.pc	as more precise	optics	to improve the
172	paiss1991.pc	separate the viewing	optics	from the nitrogen-cooled
178	pavelyev2008	Fabrication of micro	optics	on coreless fiber
190	shiryaev2013	telecommunication and non-linear	optics	systems are discussed.
192	shiryaev2013	for mid-IR fiber	optics	includes a wide
193	shiryaev2013	devices of nonlinear	optics	the optical fibers
194	shiryaev2013	glasses for fiber	optics	To prepare the

Рис. 1. Список конкордансов для ключевого слова «optics» (оптика)

Помимо этого, у программного пакета #LancsBox есть еще целый ряд функций (к сожалению, в рамках данной статьи нет возможности рассмотреть их подробно), например, поиск коллокатов для заданного ключевого слова. Результаты поиска могут быть представлены как в виде списка, так и в виде элементов визуализации, которые рассмотрены ниже.

Третье направление использования технологий цифровой лингвистики в преподавании языков — это создание коллокационных сетей, т.е. элементов визуализации языка наподобие *mind maps*, которые помогают студентам лучше запомнить коллокационные структуры и связи между ними. Это инновация разработчиков #LancsBox, и по данной теме есть только несколько статей, которые показывают общие возможности и перспективы использования данного инструмента [16]. В программе этот процесс реализуется с помощью инструмента GraphColl. На рисунке 2 приведен пример для оптики — коллокационная сеть, состоящая из двух узлов (ключевых слов) *fiber/fibre* (волокно, волоконная) и *bundle* (сборка).

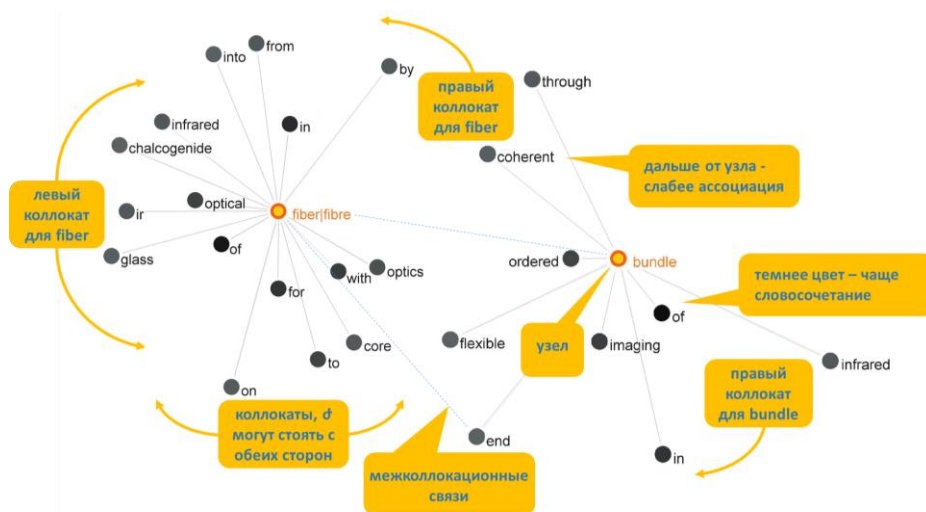


Рис. 2. Коллокационная сеть из двух узлов fiber/fibre (волокно, волоконная) – bundle (сборка)

Такие сети являются визуализациями высокочастотных коллокаций и коллигаций, с идентификацией общих коллокатов для связанных узлов (последние показаны синей пунктирной линией на рисунке). Кроме того, из коллокационных сетей могут быть найдены следующие важные параметры: 1) сила коллокации (выражается длиной линии между узлом и коллокатами – чем ближе коллокат к узлу, тем сильнее ассоциация узла и коллоката); 2) частота коллокации (определяется интенсивностью цвета коллоката – чем темнее оттенок цвета, тем чаще встречается словосочетание); 3) позиция коллоката в тексте (на графике отражается его положением вокруг узла – одни коллокаты появляются преимущественно слева от узла, другие – справа, в то время как некоторые появляются как слева, так и справа и поэтому расположены в средней позиции на графике).

Заключение

В статье проведен обзор примеров использования инструментов цифровой лингвистики в целях разработки высокоэффективных образовательных продуктов в сфере академического английского и английского для конкретных целей. Выявлено три основных пути внедрения технологий цифровой лингвистики в преподавание английского языка:

1. составление учебно-методических материалов традиционных форматов на основе передового корпусного подхода;
2. проведение студентами непосредственных исследований текстовых корпусов;
3. создание инновационных элементов визуализации языка – коллокационных сетей. На основе проведенного обзора, обоснована актуальность развития каждого из вышеперечисленных направлений для преподавания английского языка в области оптики.

Анализ данной литературы позволил обрисовать возможности, предоставляемые инструментами цифровой лингвистики для повышения качества преподавания английского языка как иностранного. Дальнейшие задачи: построение корпуса научных статей из области оптики, пригодного для проведения различных лингвистических исследований, составление учебно-методических пособий ESP по оптике, и, далее, разработка передовой программы обучения, которая бы способствовала повышению энтузиазма студентов к самообучению и, как следствие, при одной и той же контактной нагрузке, приводила бы к значительному повышению их успеваемости.

Список использованной литературы

- 1 Öksüz D., Brezina V., Rebuschat P. Collocational Processing in L1 and L2: The Effects of Word Frequency, Collocational Frequency, and Association // Language Learning. 2020 (in print).
- 2 Gablasova D., Brezina V., McEnery T. Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence // Language Learning. 2017. Vol. 67. P. 155–179.
- 3 Nekrasova-Beker T., Becker A., Sharpe A. Identifying and teaching target vocabulary in an ESP course // TESOL Journal. 2019. Vol. 10. № 1. P. e00365.

4 Brezina V., Weill-Tessier P., McEnery A. #LancsBox v. 5.x. [software] // Lancaster University corpus toolbox. Lancaster, 2020. URL: <http://corpora.lancs.ac.uk/lancsbox> (дата обращения 22.09.2020).

5 Corino E. Onesti C. Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning // *Frontiers in Education*. 2019. Vol. 4. № 7.

6 Vasheghani Farahani M., Amiri Z. The impact of teaching specialized terminology on translation performance: A corpus-based inquiry on law texts translation from English into Persian // *Journal of Applied Research in Higher Education*. 2019. Vol. 11. № 3. P. 506–521.

7 Sánchez Ramos M.M. Teaching English for medical translation: A corpus-based approach // *Iranian Journal of Language Teaching Research*. 2020. Vol. 8. № 2. P. 25–40.

8 Wang S., Zeng X. F. Effect of English Corpus on Reform of College English Teaching and the Improvement of Students' Vocabulary Competence // *Kuram ve Uygulamada Egitim Bilimleri*. 2018. Vol. 18. № 6. P. 3493–3499.

9 Louvigne S., Shi J., S. Sharmin A. Corpus-based Analysis of the Scientific RA Genre and RA Introduction // *International Conference on Advanced Mechatronic Systems, ICAMechS*. 2014. Vol. 0. № 6911636. P. 123–127.

10 Louvigne S., Jie S. Data-Driven Analysis of the Development of Linguistic Features in Research Articles on Optics // *International Conference on Advanced Mechatronic Systems, ICAMechS*. 2016. Vol. 0. № 7813502. P. 516–520.

11 Mapelli V. VERBA Polytechnic and Plurilingual Terminological Database – B – MP Optics. ISLRN: 642-374-718-061-9 // *ELRA Catalogue of Language Resources* – URL: <http://catalog.elra.info/en-us/repository/browse/ELRA-T0126/> (дата обращения 24.09.2020).

12 Evert S. Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook*. Berlin, Germany: Mouton de Gruyter, 2008. P. 1212–1248.

13 It-ngam T., Phoocharoensil S. The development of science academic word list // Indonesian Journal of Applied Linguistics. 2019. Vol. 8. № 3. P. 657–667.

14 Simbuka S., Hamied F. A., Sundayana W., Kwary D. A. A corpus-based study on the technical vocabulary of Islamic religious studies // Teflin Journal. 2019. Vol 30. № 1. P. 47–71.

15 Talalakina E., Stukal D., Kamrotov M. Developing and Validating an Academic Vocabulary List in Russian: A Computational Approach // Modern Language Journal. 2020. Vol. 104. № 3. P. 618–646.

16 Brezina V., McEnery T., Wattam S. Collocations in context: A new perspective on collocation networks // International Journal of Corpus Linguistics. 2015. Vol. 20. P. 139–173.

Кураженкова Ольга Сергеевна

ФГБОУ ВО «Курганский государственный университет»,
студентка кафедры «Безопасность информационных и
автоматизированных систем»,
ggglolya@gmail.com, Курган, Россия

Филонова Олина Игоревна

ФГБОУ ВО «Курганский государственный университет»,
канд. юр. наук, канд. ист. наук, доцент кафедры
«Безопасность информационных и автоматизированных систем»,
filonova2006@mail.ru, Курган, Россия

Человечкова Анна Владимировна

ФГБОУ ВО «Курганский государственный университет»,
старший преподаватель кафедры
«Безопасность информационных и автоматизированных систем»,
chelovechkova_2011@mail.ru, Курган, Россия

ПРОБЛЕМЫ ЗАЩИТЫ АВТОРСКОГО ПРАВА В СЕТИ ИНТЕРНЕТ

УДК 347.78:004.738.5

Аннотация. Сеть Интернет в наше время – основной способ для авторов произведений поделиться результатом своих работ, поэтому вопрос о защите их авторских прав является весьма актуальным. В данной статье проанализированы проблемные вопросы российского законодательства, призванного ограничить нелегальное использование контента, и предложены возможные пути решения.

Ключевые слова: авторское право, пиратство, плагиат, нарушение авторских прав